Using Job Vacancy Ads to Study Long-Run Occupational Change

Enghin Atalay, Sebastian Sotelo, and Daniel Tannenbaum

14th November 2017

1 Introduction

The U.S. labor market has experienced dramatic changes over the last several decades, including a steep rise in wage inequality, and shifts in the distribution of jobs over time and across geographic regions (Acemoglu and Autor 2011; Diamond, 2016). These changes have sparked a growing body of research into occupations: how the tasks that workers perform, and the rewards to workers' skills applied to different tasks have changed. Recent work has found, for example, that occupations intensive in routine and manual tasks have shrunk as a share of U.S. employment, partly due to computerization Autor, Levy, and Murnane (2003), import competition from China (Autor, Dorn, and Hanson (2013)), and increased offshorability of certain types of tasks (**?**).

One of the difficulties of studying the evolving structure of occupations, however, has been the limited availability of data measuring the skill requirements and task content of jobs. The absence of data is particular severe for researchers looking to study occupations' evolution over time and across geographies. Much of the work in this area relies on a dataset produced by the Department of Labor, called O*NET, which is based on surveys of workers, and measures occupations' skill requirements and task content. While O*NET is a valuable resource, it has important limitations: first, it is infrequently and irregularly updated, and hence can at best provide a static view of occupations, making it ill-suited for studying changes over time within occupations.[1] Second, it is hampered by the common limits of survey data, including selective non-response and recall bias.

To overcome these limitations, we have assembled a new data set of job vacancies, based on two sources: the first is a large collection of published newspaper job ads between 1940 and 2000, and the second is the near universe of online job vacancies posted between 2011-2017. With this data set, we have constructed and continue to improve a new data set of occupational tasks, technologies, and skills.[2]

[1]Papers in this literature have largely assumed that the tasks performed by each occupation are fixed through time when studying occupational changes. The predecessor of O*NET, the Dictionary of Occupational Titles (DOT), in principle allows researchers to study occupations at three points in time – 1939, 1977, and 1991 – but these data have serious limitations for longitudinal analysis, as we discuss below.

[2]This data collection has been funded through generous support from the Washington Center for Equitable Growth.

We use new computational tools developed in the field of machine learning to assemble and study this new and potentially transformational data resource to study occupations. Using the text, we extract a number of job-related elements, including tasks, skills, and technology requirements. We also link the job titles to their standardized occupational codes, so that our data can be linked to widely used data sets like the Census, and can be useful to other researchers.

Job vacancy text offers several distinct advantages over survey-based measures such as O*NET: First, it is collected in the field, where firms define the jobs and the characteristics of workers they are seeking. Second, it can be studied at high frequencies such as annual or even monthly intervals, and can be extended decades into the past as well as into the future, through the analysis of online job ads. Third, the text can be used to study occupational specialization – the number and variation of tasks within jobs, which permits an analysis of whether occupations are moving "farther apart" in their task content, and how tasks are distributed across geographic areas.

Research questions

This proposal contains several individual research projects that are at different stages of completion. What connects these projects is that they all bring our new source of data to bear on questions relating to long-run occupational change in the United States.

The research questions we address in these projects include:

- How has the task content of occupations changed over 1960-2000, and what are the implications of these changes for inequality?
- How has the arrival of Information and Communication Technologies (ICTs) transformed the tasks that workers perform, the occupations that workers sort into, and the returns to workers' different skill profiles?
- What are the patterns of specialization in tasks across local labor markets, and what are the implications for productivity and wages?

Contributions to the Russell Sage Foundation's Core Programs

This proposed research agenda is uniquely suited for the Russell Sage Foundation's initiative in Computational Social Science and its Core Program in Future of Work. We incorporate new computational tools from machine learning and natural language processing for the purposes of studying occupational change. Our approach is to extract rich data from large volumes of text appearing in job classifieds, a largely untapped source of data. Our conceptual framework embraces the task approach for labor market analysis, which represents the field's workhorse model for studying the future of work. The tools we use here can be applied to online job vacancy data as it is collected; hence our work can serve as a foundation for real-time analysis of the occupational landscape.

Our computational contribution relies on two types of tools. First, we apply Latent Dirichlet Allocation (LDA) techniques to distinguish job ads from other types of ads in the newspaper text. Second, at different points we adopt a machine learning tool called the continuous bag of words (CBOW), which analyzes large quantities of text to identify synonyms. We use the CBOW model tool to append standardized occupation codes to job ads, and to identify synonyms for task classification. The CBOW approach offers advantages over more naïve approaches, which we discuss below; this tool has had limited use in the social sciences but we believe it could be transformative for a variety of applications in which researchers are concerned about sensitivity of the analysis to diction.

Consistent with the RSF's recognition that "... the increasing cost of surveys raise questions regarding their long-run sustainability" and that "research on many topics in the social sciences is hampered by data of insufficient scale and quality..." we construct a new data set (built using data from the field) that complements O*NET, an existing survey data set.[3] The first major element of this data set are historical job vacancies from three major metropolitan newspapers – the New York Times, the Wall Street Journal and the Boston Globe – from which we retrieve more than 9 million individual vacancies for a period spanning 1940-2000. We have purchased this data from ProQuest. The second element is a large collection of online job postings, covering the period 2011-2017 for a large set of major metropolitan areas in the US. We have purchased this data set from Economic Modeling Specialists International (EMSI). Both data purchases were generously funded by a grant from the Washington Center for Equitable Growth.

This research also directly contributes to the Core Program in Future of Work. First, it provides new data and empirical evidence on the changing landscape of occupations in the U.S. Second, it studies the role of technology in shifting the occupational tasks in the U.S. The distinct advantage of our project is to directly measure technological change and technological exposure, which allows us to examine both skill-technology complementarities and technology adoption over the dramatic period of technical change experienced over the final decades of the twentieth century. The policy relevance of our work is significant: better measurement of the rapidly changing technological components of particular occupations, and the effect of these changes on other occupations, is informative both for workers making forward-looking investments in skills and for government policies aiming to promote a trained workforce.

Proposal outline

The rest of the proposal is organized as follows. Section 2 provides further description of our data sets and details the core methods of analysis that are common across the projects in the proposal. Section 3 describes key aspects of the individual projects contained in the proposal, which highlight the potential of our data set and analytic techniques. For these projects, we also provide details on the tasks that we plan to perform to bring them to completion (Section 4). In Sections 5 through 7, we lay out our work plan, justify our budget, list our expected outputs, and provide a biography of the project's team members.

[3] Our proposal also complements a recently funded Russell Sage Foundation project (Hout and Grousky, 2016). In the Hout and Grousky project, a methodological hurdle is the ability to categorize occupations according to the job titles. We describe our approach in the next section for doing so, and we hope our methods and source code can be useful to Hout and Grousky as well as other researchers facing this challenge.

2 Data and Computational Methods

In this section, we provide further detail on some of the core methods we use that are shared across all the projects included in this proposal.[4] In particular, we discuss the construction of our structured database of occupational characteristics. Once we describe these procedures, we present some simple descriptive statistics from our constructed data set. In doing so, we aim to show not only that we are aware of some of the technical and statistical difficulties pointed out to us by the reviewers, but also to show convincingly that we already solved some of those problems. Our resulting measurements, while rich and new, are plausible and echo previous findings in the literature.

As we have mentioned above, we use two primary data sets: (i) text files purchased from ProQuest, which include the full digitized content of the New York Times, Wall Street Journal, and Boston Globe; (ii) text files purchased from Economic Modeling Specialists International (EMSI), which include the text content of the near-universe of online job ads, posted between October 2011 and March 2017. Each ad contains a job title and the job description and skill requirements.

The newspaper data, because the text files were produced with an OCR software and do not come with pre-defined SOC codes, require a substantial cleaning effort that we explain below in subsections 3.1 and 3.2. In subsection 3.3, we discuss the extraction information about a job's skills and tasks from the text, by grouping sets of words according to their meaning. This procedure applies to both newspaper and online vacancies alike. We close this section discussing some preliminary findings on long-run trends in task contents and addressing the question of representativeness of newspaper data.

2.1 Cleaning the newspaper data

The newspaper data are stored as raw text files, which were produced by ProQuest using an Optical Character Recognition (OCR) technology applied to the images of the original newspapers. The raw text files allow us to isolate the subset of text that come from advertisements, but they do not directly identify which ads are for jobs and which are for other purposes, such as sales of commercial products or real estate. Hence, we apply a machine learning algorithm – a Latent Dirichlet Allocation (LDA) model – to determine which pages of advertisements are job ads. An LDA model classifies documents to one of a number of topics. In these models, the probability that different words appear in a document depends on the hidden topic of that document. LDA model estimation gives us the probability that a page belongs to a given topic, conditional on observing its text content. In our context, the LDA model is used to distinguish pages of job ads (one of the model's topics) from other groups of ads.

A second step of data cleaning is to identify where a job ad ends and another ad begins. Figure 1 presents a portion of a page of job ads. This snippet of text refers to three job ads, first for a Software Engineer position, then a Senior Systems Engineer position, and finally for a Software Engineer position. Within this page of ads, we determine the boundaries of each individual advertisement (where, e.g., the Software Engineer ad ends and the Senior

[4] In this section we borrow substantially from our draft Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2017).

Systems Engineer ad begins). In particular, we use a rule that relies on certain common patterns across ads and relies on the presence of (i) street addresses, (ii) certain phrases (e.g., "send [...] resume" or "equal opportunity employer"), and (iii) the formatting of job titles; specifically we look for lines of text that are either all capitalized or in which only one or two words appear. Within these lines, we search for words which appear in O*NET's "Sample of Reported Job Titles." If there is a match, we record the contents of the lines as the ad's job title.

2.2 Assigning SOC codes to job ad data

In order for us to make our processed data as useful as possible to other researchers, and to permit linking the processed data to other data sets, we attach standard occupation classifications to each ad, using the ad's job title. For example, vacancy postings for registered nurses will be advertised using various job titles, including "rn," "registered nurse," or "staff nurse." These job titles should all map to the same occupation — 291141 using the BLS Standard Occupational Classification (SOC) system.

From our list of job titles, we apply two complementary methods to attach SOC codes to ads. First, for the most common 1000 job titles from our Boston Globe, New York Times, and Wall Street Journal job ads, we map the titles to their SOC codes using www.onetsocautocoder.com. These top 1000 job titles cover 2.2 of the 6.6 million ads from the previous subsection's processed newspaper data. For the remaining ads, we apply a machine learning algorithm called a continuous bag of words model (CBOW), in combination with the EMSI online ads data, to attempt to identify the ad's SOC code. The CBOW model is described in detail in Appendix C of Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2017), but the essential idea is to identify synonyms for words or phrases based on where they appear relative to surrounding words (in text corpora). For example, to the extent that "nurse," "rn," and "staff nurse" all tend to appear next to words like "patient," "care," or "blood" one would conclude that "rn" and "nurse" have similar meanings to one another. The CBOW model is useful since the EMSI data has a large correspondence between job titles and SOC codes.

We use the online job postings from two of these months, January 2012 and January 2016, plus all of the text from our newspaper data to construct our CBOW model. We use this model at different points in the paper to find synonyms for words or phrases. At this stage, our CBOW model allows us to identify the closest job title in the EMSI data set for each job title in our newspaper text. Since job titles in the EMSI data set have an associated SOC code, we can obtain the SOC code for any job title in our newspaper text. We retrieve these SOC codes on all of the job titles that appear at least twice in our newspaper text. In combination with the first approach, we have been able to assign SOC codes to 4.1 million job ads.

2.3 Extracting tasks and skill requirements from each ad

Our next step is to consolidate the information in our vacancy postings into a small number of economically meaningful task categories. Our main classification follows that of Spitz-Oener (2006) who, in her study of the changing task content of German occupations, groups survey questionnaire responses into five categories: nonroutine analytic, nonroutine interactive, nonroutine manual, routine cognitive, and routine analytic. We begin with Spitz-Oener's (2006) list of words corresponding to each of her five tasks.

We then augment each task category's corresponding set of words to include their synonyms as well, where synonyms are defined through the lens of the CBOW model. The idea to let the text define task categories, through a statistical model-based approach, which achieves two aims: first, it limits the amount of discretion given to the researchers in categorizing words into tasks.[5] By using the CBOW model we are in some sense allowing the text to define task categories by grouping words based on statistical likelihood of appearing in the same context within text corpora.

We emphasize that our CBOW model-based approach is well suited for limiting our data's sensitivity to trends in language -- either word usage or meaning. To give a concrete example, suppose the word "researching" falls out of favor among employers, who instead use the word "investigating." Since one of the words corresponding to the nonroutine analytic task is researching, a naive approach of searching for the word "researching" would lead to a researcher to erroneously conclude that nonroutine analytic tasks are declining in importance in the labor market. But the continuous bag of words model correctly identifies synonyms of researching, including the words "interpreting", "investigating", and "reviewing". Hence, our CBOW-based task measures would count investigating as a nonroutine analytic task and would not be subject to the error due to changes in language. By including the union of synonyms for the nonroutine analytic task, we limit the sensitivity of our analysis to variations in diction over time.

In addition to the Spitz-Oener based classification, we connect to other results in the literature by adopting the classifications in Deming and Kahn (2017), Firpo, Fortin, and Lemieux (2014), and the Occupational Information Network (O*NET). In particular, we demonstrate that the measures that result from our mapping correlate, across occupations, with O*NET's measures of the importance of different tasks.[6]

2.4 Descriptive statistics

Table 1 lists the top occupations in our data set. The first two columns list common job titles among the 6.6 million total vacancy postings, while the last four columns present the top SOC codes. Across the universe of occupations, our newspaper data represents a broad swath of Management, Business, Computer, Engineering, Life and Physical Science, Healthcare, Sales, and Administrative Support occupations, but it under-represents Construction occupations and occupations related to the production and transportation of goods. [7]

Table 2 presents, for each task in Spitz-Oener (2006)'s classification, the most task-intensive occupations. Occupations that are intensive in nonroutine analytic tasks are con-

[5]A common critique of research into occupational tasks is the amount of discretion left to researchers over task categorization, which could be chosen to suit the researchers' priors (Autor, 2013)

[6]See Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2017) for further detail on these validation exercises.

[7]See Appendix A of Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2017), as well as our answers to the reviewers, for an analysis of the representativeness of our newspaper data relative to the decennial census and CPS. (These latter data sets we access via **?**)

centrated in Architectural and Engineering occupations (SOC codes beginning with "17") and Life, Physical, and Social Science Occupations (SOC codes beginning with "19"). Management ("11") and Sales ("41") occupations mention nonroutine interactive tasks frequently, while customer-service and maintenance related occupations (in a variety of SOC groups) have high nonroutine manual task content. Routine cognitive and routine manual task related words are mentioned frequently in advertisements for clerical and production-related positions.

Perhaps the most attractive features of the data are the large sample size, measured at high frequencies, over a long time horizon. We are able to observe the evolution of occupations over time, including changes in tasks (highlighted in Section 3.1), which is currently unavailable to researchers and opens up a pathway for research into the U.S. occupational transformation. An additional feature the data is that it is collected in the "field", meaning that firms define, using their own words, the jobs they seek to fill. Unlike survey data, our data are not subject to recall-bias, selective nonresponse, or confusion about survey questions, a well-known weakness of O*NET (Autor, 2013).[8] A third key strength is that job ads represent the frontier of occupational change, making the data particularly useful for tracking the evolution of the U.S. labor market.

2.5 Representativeness

An immediate question about this new data source is to what extent job ads appearing in newspapers are representative of overall job vacancies in the U.S.? In particular, we might be concerned that the trends we identify in the newspaper data might not be representative of vacancies in the rest of the country. We study this potential problem from three different angles, and conclude that our sample can be informative of the rest of the population. We report the results of this inquiry in Appendices A and B of Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2017).

First, we look for trends in job search behavior over time: i.e. whether workers are more likely to search for jobs using ads compared to other sources over time, and whether there are particular types of workers exhibit trends (e.g. high and low educated, or workers with experience in high or low routine or nonroutine occupations). We conclude that this does not appear to occur. Second, we compare shares of employment across occupations in the decennial census with shares of vacancies in the newspaper data, with the idea that a stable relation between these shares over time is consistent with similar trends in task content. Here we conclude that the relationship has not changed substantially. Finally, for the ads where we can recover information on educational requirements, we compare those with the actual educational attainment of workers observed in the decenial census. Across occupations, we find positive correlations in whether the occupation requires a colleges chool degree. We are planning to conduct further tests, which we detail below in subsection 3.1.

[8]Throughout this paper, we interpret the words in job ads as accurate descriptions of the positions the firms seek to fill. We cannot measure the extent to which firms may misrepresent or perhaps euphemize the tasks of the job to attract workers. A similar consideration, however, is also relevant for survey-based measures of tasks, where respondents may inflate or deflate their job's true task content in their responses (Autor, 2013). Our analysis is unaffected by level differences job descriptions' accuracy, but would be affected by trends in accuracy.

Importantly, in principle nothing hinders the application of our approach and publicly available source code to other newspapers or other job adverstisement collections. In practice, however, we faced two major obstacles. For one thing, our budget only allowed for the purchase of a few metropolitan newspapers. For another, ProQuest, our data provider, could not provide us with rights to mine the data of other newspapers we could have potentially purchased. Moreover, as we explain below, the online job ads have information on geographic location, which we wish to explore both to provide further checks on representativeness and to study variation across cities in task compositions.

3 Four Projects on Occupation Characteristics

We now outline projects for future study as part of our overall research agenda, highlighting research questions that exploit the strengths of our new data set. While these projects are preliminary, they demonstrate the substantial promise of this data for informing the Future of Work.

3.1 Project 1: The Evolving Occupational Structure

One of the striking findings of labor economics in the past two decades has been the restructuring of U.S. employment across occupations: the employment share of occupations intensive in routine tasks has shrunk, while the share of occupations emphasizing nonroutine tasks, as well as social and cognitive skills, has grown (Autor, Levy, and Murnane, 2003; Autor, Katz, and Kearney, 2005; Autor and Dorn, 2013).[9]

The evidence thus far points to the decline of employment in U.S. occupations intensive in routine tasks, but is largely silent on whether the content of occupations themselves has changed. Meanwhile, case studies of individual occupations have found considerable changes. For example, a report by the National Research Council (1999) suggests that managerial occupations over the second half of the twentieth century increasingly have emphasized team management, coaching skills and tasks, and interaction with customers. The rise in these skills and tasks were accompanied by a decline of managerial tasks related to direct control of subordinates. These findings raise the question of whether managerial occupations are unique in experiencing changes in tasks, or if comparable changes have occurred elsewhere.

Moreover, to the extent that the tasks that workers perform on the job have shifted, what are the implications for the distribution of earnings? In this project, we show that substantial changes in task composition did occur within occupations since 1960. Moreover, we find that these within-occupation changes in task content account for much of the observed increase in earnings inequality. Using our new data set, we show that words related to interactive tasks have been increasing in frequency, while words related to routine tasks (especially routine manual tasks) have declined in frequency between 1960 and 2000. For instance, we find that the frequency of words related to routine cognitive tasks has declined by almost half from the early 1960s to the late 1990s, from 1.00 mentions per thousand job ad words to

[9]This section draws heavily on our ongoing work Atalay, Phongthiengtham, Sotelo, and Tannenbaum (2017). This project and Project 2 are joint work with Phai Phongthiengtham, who is a graduate student in the University of Wisconsin-Madison Department of Economics.

0.54 mentions per thousand words. The frequency of routine manual tasks has declined even more starkly. On the other hand, the frequency of words related to nonroutine interactive tasks has increased by more than 54 percent, from 3.19 to 5.48 mentions per thousand of job ad words. Importantly, we find that a large fraction of the aggregate changes in both nonroutine and routine task related words has occurred within occupations, rather than through changes in the occupations' employment shares.

Having documented these patterns, we provide two alternative quantifications of the relation between changes in the task content of occupations and the rise of inequality in the U.S. The first one relies on decomposition methods applied to the earnings distribution; the second, on an equilibrium model of worker occupational choices and comparative advantage.

In our first quantification, we incorporate our occupational measures into Firpo, Fortin, and Lemieux (2014)'s methodology for decomposing changes in the wage distribution across points in time. Using these methods, we break down changes in the distribution of earnings over time into changes in the attributes of workers and their occupations (the "composition" effect), as well as changes in the implicit prices of those observable characteristics (the "wage structure" effect). Next, we further break down changes in the distribution of earnings into the contributions of observable characteristics (e.g., the contribution of changes in educational attainment and of changes in the returns to education).

Our results suggest that, relative to the upper tail of the income distribution, tasks that were valued highly in low-wage jobs (in particular manual tasks) declined in importance. These task changes, however, come more from changes within the occupations themselves than from shifts of employment across occupations. In the decomposition, changes in occupational task content account, through composition effects, for a 32 log point increase in 90-10 male earnings inequality. Wage structure effects — changes in tasks' implict prices — account for a 6 percentage point decrease in 90-10 inequality. Among the 32 percentage points that are due to composition effects, routine manual and nonroutine interactive tasks respectively account for a 24 and 8 log point increase in inequality. (The contribution of changes in routine cognitive, nonroutine analytic, and nonroutine manual tasks is more modest.) If we instead ignore the evolution of occupational characteristics across time, and keep the task measures fixed, we can only account for a total of 3 percentage points of the observed increase in 90-10 inequality. Hence occupational measures of task intensity account for a large component of the increase in earnings inequality between 1960 and 2000, but only if task measures are allowed to vary across time.

In our second quantification, which we view as complementary to the earnings distribution decomposition, we construct a general equilibrium model of occupational choice. In our model, individual occupations are represented as a bundle of tasks. Workers' skills govern their abilities to perform each of the individual tasks in their occupation, and give rise to comparative advantage. These skill levels are functions of workers' observable characteristics — like gender, education, and experience — but also contain an idiosyncratic component.

Based on their skill levels and the demand for tasks within each occupation, workers select into the occupation with the highest payoff. We estimate the skills by which each demographic group performs different groups of tasks, combining information on demographic groups' wages and occupation choices. Using our estimated model, we calculate that changes in the relative demands for tasks have led to a 25 log point increase in inequality among men, and a 22 log point increase in inequality among women and men. The intuition is that less educated workers both tend to be at the bottom of the wage distribution and have a comparative advantage in manual tasks, while the demand for the latter has declined.

Both the statistical decompositions and our model-based counterfactual exercises have their own advantages. In spite of their differences, however, both methods indicate that a relative decline in the demand for routine tasks has substantially increased earnings inequality between 1960 and 2000. Moreover, both of these exercises require information on changes in occupations' task content, something that our data set is suited to measure.

Project 1 builds on two literatures, the first of which examines the causes and consequences of the evolution of occupations. Autor, Levy, and Murnane (2003) and Acemoglu and Autor (2011) develop the hypothesis that technological advances have reduced the demand for routine tasks, which, in turn, has led to a reduction in the wages of low- and middle-skill workers. More recently, Firpo, Fortin, and Lemieux (2014) decompose changes in the distribution of wages into the contribution of occupational characteristics and other factors (including de-unionization, changes in minimum wage, and changes in worker demographics). Deming (2017) documents that employment and wage growth has been confined to occupations that are intensive in both social and cognitive skills. Michaels, Rauch, and Redding (2016) extend Autor, Levy, and Murnane (2003) to study changes in employment shares by task content over a longer time horizon. They use a methodology that is related to ours, using the verbs from the Dictionary of Occupational Titles' occupational descriptions and their thesaurus-based meanings. Burstein, Morales, and Vogel (2015) quantify the impact of the adoption of computers on between-group wage inequality. Because commonly available data sets measuring occupational characteristics are infrequently and irregularly updated, in most of these papers occupational characteristics are fixed over time. One paper that focuses on within-occupation changes is Spitz-Oener (2006), which uses survey data from four waves of German workers to track skill changes within and between occupations, from the late 1970s to the late 1990s. One of our contributions is to study these changes in the U.S. context, extending the Spitz-Oener (2006) task classifications, and examining the implications of the observed changes in tasks for the wage distribution.

Our paper contributes more generally to the measurement of time-varying task content of occupations by using newspaper job ads, a rich and largely untapped source of data, and which are complementary to existing data sources currently used to study the evolution of the U.S. labor market.

A second related literature uses the text from online help wanted ads to study the labor market: how firms and workers match with one another, how firms differ in their job requirements, and how skill requirements have changed since the beginning of the Great Recession. Using data from CareerBuilder, Marinescu and Wolthoff (2016) document substantial variation in job ads' skill requirements and stated salaries within narrowly-defined occupation codes. Moreover, the within-occupation variation is to a large extent explained by the vacancy posting's job title. Using online job ads, Hershbein and Kahn (2016) and Modestino, Shoag, and Ballance (2016) argue that jobs' skill requirements have increased during the post-Great Recession period; Deming and Kahn (2017) find that firms that post ads with a high frequency of words related to social and cognitive skills have higher labor productivity and pay higher wages. By bringing in newspaper data, we contribute to this literature with measurement of long-run changes in tasks and skills in the labor market, which previously could not be measured.

Preliminary findings. We briefly present three vignettes of occupational change that highlight how occupations have evolved over our sample period – an evolution that prior to our data was not possible to observe.

In Figure 2, we present two separate task measures for managerial occupations (in thick, dashed lines) and all occupations (in thin, solid lines). Between 1960 and 2000, the frequency of words related to nonroutine interactive tasks in managerial occupations increased by approximately one-third, from 7.5 to 10.0 mentions per 1000 ad words. This measured trend aligns with the National Research Council (1999)'s characterization of the changing nature of managerial work. Summarizing the contemporaneous literature, the National Research Council (1999) write that trends in managerial work involve "the growing importance of skills in dealing with organizations and people external to the firm, ... the requirement that [managers] 'coach'... and facilitate relations between workers." (pp. 137-138) Motivated by this characterization, we also plot trends in the mentions of four O*NET work activities: Working with the Public (O*NET Element 4.A.4.a.3), Establishing and Maintaining Relationships (4.A.4.a.4), Building Teams (4.A.4.b.2), and Coaching (4.A.4.b.5). Mentions of these four activities increased by more than 50 percent in managerial occupations, increasing more quickly than for the workforce as a whole. In sum, while interactive tasks have always been a requirement for managerial occupations, the importance of such tasks has widened since 1960.

Second, Figure 3 contrasts trends in task measures for office clerks compared to all occupations. While mentions of routine cognitive tasks has decreased for both office clerks and more generally for all occupations, the drop off has been more pronounced in clerical positions. Concurrently, the frequency of nonroutine interactive task-related words has increased in clerical ads roughly at the same pace as in other occupations.

Finally, compared to the beginning of the sample period, the frequency of routine manual tasks has declined considerably, particularly in non-supervisory production occupations. Figure 4 presents these trends along with changes in the frequency of nonroutine analytic task related words. Nonroutine analytic task keywords have been increasingly mentioned in job ads for production workers. The trends in keyword frequencies in this figure are consistent with case studies of manufacturers' adoption of new information technologies (e.g., Bartel, Ichniowski, and Shaw, 2007). These new technologies substitute for workers who were previously performing routine manual tasks. The surviving production workers are those who have high levels of technical and problem-solving skills.

Planning ahead on Project 1

- We wish to conduct additional exercises to demonstrate representativeness of NYC and Boston metro newspapers for job vacancies. The first is to compare the task content of occupations in Boston and NYC to the rest of US using the EMSI data (which has complete geographic coverage for post-2010).

- We would like to assess the usefulness of the Dictionary of Occupational Titles (DOT) for observing within-occupational changes. In principle, the 1939, 1977, and 1991 DOT could be combined to create 3 data points of occupations that could be linked to study occupational change, although these data have been criticized for status quo bias (Miller et al., 1980).

- Describe our job title to SOC match rates over time and across labor markets.
- Provide additional background information on newspaper job classifieds. Specifically (1) trends in circulation of the newspaper readership (2) demographic trends in readership (e.g. education) (3) does our analysis period precede the decline of newspaper readership? (3) have ad prices in these newspapers changed a lot over time or have they been relatively stable?
- Provide further evidence on whether the biggest changes in task content within occupations are for those that were most intensive in those tasks.

3.2 Project 2: New Technologies and the Labor Market

Enabled by increasingly powerful computers and the proliferation of new, ever more capable software, the fraction of workers' time spent using information and communication technologies (ICTs) has increased considerably over the last half century.[10] Nordhaus (2007) estimates that, between 1960 and 1999, the total cost of a standardized set of computations fell by between 30 and 75 percent annually, a rapid rate of change that far outpaced earlier historical periods. In this project, we quantify the impact of 40 individual and recognizable ICTs on the aggregate demand for routine and nonroutine tasks, on the allocation of workers across occupations, and on earnings inequality.

We begin this analysis by first extracting job-level information about the adoption of 40 ICTs across occupations and years in the job ads, as measured by their appearance in the text of job postings. The technologies we study constitute a wide set, ranging from office software (including Lotus 123, Word Perfect, Microsoft Word, Excel, PowerPoint), enterprise programming languages (Electronic Data Processing, Sybase), general-purpose programming languages (COBOL, Fortran, Java), and hardware (UNIVAC, IBM 360, IBM 370), among others. With this data set, we document that, for the most part, job ads that mention a new technology tend to also mention nonroutine analytic tasks more frequently, while mentioning other tasks less frequently, this provides preliminary evidence that new technologies are complementary with particular tasks. An important exception is office software, which is more likely to appear alongside words associated with nonroutine interactive tasks.

Since our data set includes a wide range of occupations and technologies, we can speak directly to the macroeconomic implications of changes in ICT prices while maintaining a detailed analysis of individual occupations. Informed by our micro estimates on the relationship between the tasks that workers perform and the technologies they use on the job, we build a quantitative model of occupational sorting and technology adoption. In the model, workers sort into occupations based on their comparative advantage. They also choose which ICT to adopt, if any, based on the price of each piece of technology and the technology's complementarity with the tasks involved in their occupation. Within the model, the availability of a new technology, which we model as a decline in the technology's price, alters the types of tasks workers perform in their occupation.

[10]This section draws substantially from ongoing work for a manuscript entitled "New Technologies and the Labor Market," which is co-authored with Phai Phongthiengtham.

Using our model, we explore the effect of the arrival of ICTs by studying counterfactual scenarios where groups of technologies were never introduced. One of our main findings is that new technologies result in an increase in occupations' nonroutine analytic task content (relative to other tasks). As we document in Project 1, and report in the first manuscript using this data (Atalay et al., 2017), workers with observable characteristics indicating high skill levels (experienced and highly educated workers) have a comparative advantage in producing nonroutine analytic tasks. Because new technologies increase the demand for worker-performed tasks in which high-skilled workers have a comparative advantage, the introduction of ICTs has (for the most part) led to an increase in income inequality. Unlike the other technologies in our data, Microsoft Office technologies are only weakly correlated with nonroutine analytic tasks. Concomitantly, the impact of Microsoft Office software has been to decrease the skill premium, and income inequality. However, the effects of these technologies are small.

This project relates to a rich literature in labor economics exploring the implications of technological change for skill prices and the wage distribution (Katz and Murphy (1992); Juhn, Murphy, and Pierce (1993); Berman, Bound, and Machin (1998); Krusell, Ohanian, Rios-Rull, and Violante (2000)). More recent work has argued that information technology complements high-skilled workers performing abstract tasks and substitutes for middle-skilled workers performing routine tasks (Autor, Levy, and Murnane (2003); Goos and Manning (2007); Autor, Katz, and Kearney (2005); Acemoglu and Autor (2011)).

This project adopts the task approach as well, and examines how new technologies complement (or substitute for) the types of tasks that workers of different skill groups perform, finding that ICTs tend to substitute for routine tasks (especially routine manual tasks) which are disproportionately performed by low skill workers. In turn ICTs allow high skill workers to focus on the activities in which they are the most productive, which in our model is the essence of the complementarity. One of our contributions to this literature is to measure both technological change and the task content of occupations directly, over a period of immense technological change.

Our paper relates to a second literature that measures directly the adoption of specific technologies and its effect on wages and the demand for skills. These include studies of the effect of computer adoption (e.g., Krueger (1993); Entorf and Kramarz (1998); Haisken-DeNew and Schmidt (1999); Autor, Katz, and Krueger (1998)) or the introduction of broadband internet (e.g., Brynjolfsson and Hitt (2003); Akerman, Gaarder, and Mogstad (2015)) on worker productivity and wages. Also exploiting text descriptions of occupations, Michaels, Rauch, and Redding (2016) provide evidence that, since 1880, new technologies that enhance human interaction have reshaped the spatial distribution of economic activity. Focusing on a more recent technological revolution, Burstein, Morales, and Vogel (2015) document how the diffusion of computing technologies has contributed to the rise of inequality in the U.S. Our paper builds on this literature by introducing a rich data set measuring the adoption of ICTs at the job vacancy level.

In what follows, we provide some more detail on the preliminary findings that come out of this project. Specifically, we discuss (i) measurements of technology adoption using our newspaper data, and (ii) measurements of complementarity between ICTs and tasks.

Measuring technology adoption. Table 3 lists the technologies in our sample together with information on their timing of adoption, as measured by the number of mentions in job ads, and the year the technology was introduced.[11] The columns titled "First Year" and "Last Year" list the first and last years within the 1960 to 2000 period in which the frequency of technology mentions is at least one-third of the mentions in the year when the technology is mentioned most frequently. Using this one-third cutoff, the lag between technology introduction and technology adoption (i.e. the difference between the "Introduction" and the "First Year" column) is 8 years on average. The final column lists the overall frequency of mentions, across the 6.6 million job ads in our data set, of each piece of technology.

Figure 5 plots the trends in technology mentions in our data set. Over the sample period, there is a broad increase in the frequency with which employers mention technologies, from less than 0.02 mentions per ad in the beginning of the sample to 0.20 mentions by 2000. While there is a broad increase in technology adoption rates throughout the sample, certain technologies have faded from use over time. The right panel of Figure 5 documents adoption rates for each of the 40 technologies in our sample, with seven of these highlighted. Certain technologies which were prevalent in the 1960s and 1970s — including Electronic Data Processing (EDP) and COBOL — have declined in usage. Other technologies — Word Perfect and Lotus 123 — quickly increased and then decreased in newspaper mentions.

In Figure 5, we examine the heterogeneity across occupations in their adoption rates. Here, we plot the frequency of job ads which mention each technology, across 4-digit SOC groups of four different technologies: Fortran, Unix, Word Perfect, and Microsoft Word. Each plot indicates with a vertical line the year of release of the technology to the public. These plots suggest several new facts. First, technological adoption is uneven across occupations, occurring at different times and to different degrees. For instance Fortran is quickly adopted by Computer Programmers, while the adoption by Engineers lags behind and is more limited. Second, for technologies that perform the same function, such as Word Perfect and Microsoft Word, the figures suggest dramatic substitution between technologies. Lastly, we see that office software is adopted widely across diverse occupations, whereas other types of software, such as CAD, are adopted more narrowly. Finally, between the time of release to the public and the peak of adoption, adoption rates increase first quickly and then slowly. This pattern is consistent with the S-shaped documented in the diffusion of many technologies (e.g., Griliches, 1957; Gort and Klepper, 1982). While we do not offer a theory of the pattern of adoption of new technologies for each occupation, we will exploit the time variation in adoption rates to gauge their impact on the macroeconomy.

Complementarities between task and technology adoption. Our job ads data set allows us to investigate the degree of complementarity between tasks and technologies for the adopting occupations. In our data, new technologies tend to be mentioned jointly with analytic tasks, not with nonroutine interactive, nonroutine manual, routine cognitive, or routine manual tasks. There are important exceptions, however, such as the widely adopted office software and interactive tasks.

To show this, we exploit temporal and occupational variation in the extent to which workers adopt technologies, we estimate the following equation:

[11] We obtained the year of introduction from the Wikipedia page of each technology.

$$\text{task}^{h}_{ajt} \;=\; \beta_{hk} \cdot \text{tech}_{ajkt} + f_h\left(\text{words}_{ajt}\right) + \iota_{jh} + \iota_{th} + \epsilon_{ahjkt} \tag{1}$$

In Equation 1, h refers to one of five potential routine and nonroutine task categories; tech_{ajkt} gives the number of mentions of a particular technology k in individual job ad a, published in year t for an occupation j; ι_{jh} and ι_{th} refer to occupation and year fixed effects, respectively; and $f_h\left(\text{words}_{ajt}\right)$ is a quartic polynomial controlling for the number of words in the ad, since the word count varies across ads. We run the regressions characterized by Equation 1 separately for each technology k and task h. The occupation fixed effects and year-fixed effects respectively control for occupation-specific differences in the frequency of task mentions and economy-wide trends in the tasks that workers perform unrelated to technology adoption.[12]

Figure 6 presents the estimates of β_{hk} for each task-technology pair. Within each panel, technologies are grouped according to their type, with database management systems first, then office software, networking software/hardware third, other hardware fourth, and general purpose software fifth. According to the top-left panel, the relationship between nonroutine analytic task mentions and technology mentions is increasing for database management systems, networking software/hardware, and general purpose software. Among the 40 technologies in our sample, the median effect of an additional technology-related mention is an additional 0.05 nonroutine analytic task mentions per job ad. On the other hand, technology mentions and task mentions are broadly inversely related for the other four task categories: An additional mention of a technology is associated (again, according to the median of the 40 coefficient estimates) with 0.137 fewer mentions of nonroutine interactive tasks, 0.018 fewer mentions of nonroutine manual tasks, 0.011 fewer mentions of routine cognitive tasks, and 0.017 fewer mentions of routine manual tasks.[13] But there are important exceptions to these interactions: Quark XPress, CAD, Microsoft Excel, and PowerPoint are the four technologies which are associated with an increasing frequency of nonroutine interactive task-related words. Three of the networking technologies — LAN, Novell, and TCP — are associated with increased mentions of routine cognitive task mentions.

3.2.1 Planning ahead on Project 2: incorporating feedback

- Attempt a more general method for identifying ICTs in our data set. In particular, our current set of ICTs is more exhaustive in listing software, but relatively more sparse when it comes to hardware.
- Explore more carefully the role of occupations in adopting ICTs. In our sample, a substantial part of adoption is driven by engineering related occupations. We would

[12] The identifying assumption is that variation in adoption rates is driven by changes in the relative prices of different ICTs and the extent to which ICTs interact with task content.

[13] The frequencies with which employers mention tasks — and with which our text-processing algorithm detects task-related words — differ across the five task categories. Stating our coefficients in a comparable scale, the median effect of an individual technology mention is associated with a 0.07 standard deviation increase in nonroutine analytic task mentions, and a decline in nonroutine interactive, nonroutine manual, routine cognitive, and routine manual task mentions of (respectively) 0.20, 0.06, 0.05, and 0.11 standard deviations.

like to understand better what technologies different occupations adopt as well as what has been the effect of non-engineering occupations on the aggregate.

- The estimation of our equilibrium model of working sorting uses only data from the year 2000, in an attempt to minimize the role of discrimination in labor markets. A robustness check that we'd like to perform, however, is check for the time invariance of our comparative advantage patterns by estimating the model for the year 1980 as well.

- We propose to extend the general equilibrium model to study how the adoption of technologies in one occupation spills over to occupations that are connected through industry links.

3.3 Project 3: Geography of Specialization in Tasks

One of the unique features of our online job ads data is that it covers the entire U.S., spanning urban, rural, and suburban areas. We propose to explore the geographic differences in occupational tasks, and the implications of these differences for wage inequality across space. In particular, we propose to first develop an empirical measure of occupational specialization – encompassing the number and variation of tasks – which permits an analysis of how tasks for a given occupation vary across geographic areas. Next we propose to build a quantitative model, combining features of **?** with the tractability of Hsieh, Hurst, Jones, and Klenow (2016), which will allow us to answer the following question: How much of the difference in productivity between cities is due to the difference in specialization of their workers, induced by different city size?

This project is inspired by the influential work of **?**, which introduces a theoretical model showing that when workers can make task-specific human capital investments, there are increasing returns to specialization.[14] One of the empirical predictions of their model is that increases in the number of people in a local labor market will lead to a greater degree of specialization in tasks and increases in wages. While there is an ample literature documenting the higher productivity and wages in cities (**??**), thus far there has been no comprehensive approach to measuring specialization in tasks, due a lack of available data.[15] Our data set puts us in an ideal position to conduct such a study. The contribution of this project is to assess the macroeconomic implications of specialization, and to provide a quantitative, empirically grounded, link between occupational specialization, market size, and productivity.

3.4 Project 4: Tasks and Demographic Groups

In Projects 1 and 2, we document the transformation of U.S. occupations over 1960-2000. This period coincides with key demographic changes in the workforce; in particular, the rise

[14]**?** presents a brief intellectual history of this idea, and an application to vertical integration.

[15]Prior research has studied the degree of specialization for specific occupations, such as lawyers (**?**), and a recent literature in natural sciences has shown that there is greater professional diversity in larger labor markets (**?**). A related study by **?** shows that specialists are over-represented in large French cities.

of female labor force participation.[16]

The newspaper job ads, combined with our computational tools may offer a window into gender discrimination before and after the Civil Rights Act of 1964, which banned explicit targeting of job ads to genders. We propose to examine gender-targeted job ads prior to 1964, and then to examine whether gender-targeting continued, or even increased, following the Civil Rights Act through the use of "code words." A typical example is the use of "word per minute" requirement for assistants, which might used to target female applicants. The tools of natural language processing will allow us to identify these code words, by identifying a set of words that commonly appear with gender pronouns prior to 1964; we can then examine whether there is an increase in these code words following the passage of the law. In this way we can apply natural language processing to identify patterns discrimination over the 20th century.

4 Work Plan

This section outlines the timeline for each of the projects above.

4.1 Projects 1 and 2

July 2018 - December 2018

- Anticipated revision and resubmission of Project 1 draft to a general interest or field jounal. Work on revisions to writing, descriptive statistics, and modeling.
- Project 2 is slated for publication on the July 2018 Volume of the Journal of Monetary Economics. Work leading to this point will include final revisions of modeling and estimation.
- Develop website making available public use data and codes associated with (i) CBOW-based mapping of job titles to SOC codes; (ii) CBOW-based mappings of words to tasks; (iii) Newspaper-based occupation-year data including tasks, skill requirements, and ICT technologies used on the job.

January 2019 - June 2019

- Anticipated acceptance for publication of Project 1 in a leading general interest or field journal. Final revisions to draft

[16] As part of Projects 1 and 2, we explore the role of gender and occupational sorting by task content, by estimating models of comparative advantage. Our results suggest that women tend to have a comparative advantage in nonroutine interactive tasks relative to men. One implication of this finding is that the increase in the demand for interactive tasks has helped close the gender gap in wages. A related finding is that certain technologies, such as Microsoft Excel, Powerpoint, and Word, have also helped close the gender gap because they interact positively with nonroutine anaytic tasks.

4.2 Projects 3 and 4

July 2018 - December 2018

- Descriptive analysis of EMSI online job vacancies (Project 3)
- Develop quantitative model of task specialization across cities (Project 3)
- Complete first draft of manuscript (Project 3)

January 2019 - June 2019

- Present project at conferences and seminar invitations (Project 3)
- Revise draft in response to feedback received during presentations (Project 3)
- Extend our cleaned occupation database to reach back to 1940. Our job ads data has been cleaned back to 1960 but we would like to study longer-run changes in the task content of occupations, and propose to extend our database back to 1940 using already-purchased data. This activity requires substantial research assistance. (Project 4)

July 2019 - December 2020

- Anticipated date of submission of draft to a leading general interest or field journal (Project 3)
- Complete draft first draft of manuscript (Project 4)

January 2020 - June 2020

- Revisions to draft following feedback from refereeing process (Project 3)
- Resubmission of draft for publication (Project 4)

5 Budget Narrative and Justification

University of Wisconsin – Madison Budget Justification and Scope of Work

Senior Personnel: PI Enghin Atalay, Asst. Professor (approx. 1 summer month in Year 1 and approx. 0.5 summer month in Year 2 with a base salary increase in Y2 of 3% from Y1). During the project, my primary responsibilities will be to write and present the results of the proposed research and to supervise the Project Assistant in the tasks that he/she needs to perform.

Research Assistance: Approx. a 42% academic Project Assistant for 7.5 months in Year 1 and about an 11% PA for 2 summer months in Year 2 with a base salary increase in Y2 of 3% from Y1. The PA scope of work incudes managing and executing the data cleaning process, implementing the continuous bag of word model on the text, conducting empirical analysis on the cleaned data, and assistance in writing up supporting documentation and results.

Fringe Benefits: 36% PI and 24% PA in Year 1. 37% PI and 25% PA in Year 2.

Current & Pending Support: None related to this proposal except "this proposal", $114,772 (which includes the 2 subcontracts below). 7/01/2018-6/30/2020.

University of Michigan Budget Justification and Scope of Work

Senior Personnel: Sebastian Sotelo, Asst Professor (0.75 summer month Year 1 and 0.50 summer month Year 2). My main tasks are reading and synthesizing the relevant literature, developing a new model that takes the insights of that literature and extends them to help interpret our data, and conducting data analysis and simulation according to the theory.

Fringe Benefits: The University of Michigan does not have a negotiated fringe benefit rate. Rather estimates of probable benefits expenses are used in proposal budgets; grants are then charged for actual benefits expenses incurred, in direct proportion to the applicable salaries and wages charged. Individual units are encouraged to examine their benefits patterns to determine an appropriate estimator. Currently, campus-wide benefits average 30%. Rates can be found at http://orsp.umich.edu/proposals/budgets/staff_benefits_table.html.

Current & Pending Support: None related to this proposal except "this proposal", $29,124. 7/01/2018-6/30/2020

University of Nebraska Budget Justification and Scope of Work

Senior Personnel: Daniel Tannenbaum, PI, (effort = 0.8 summer months in year 1 and 0.5 summer months in year 2) will be responsible for conducting the empirical analysis laid out in the proposal, writing articles for publication in academic journals, and communicating our results to the broader research community and other stakeholders.

Fringe Benefits: Estimated at 30% for the PI. The actual cost of benefits for each person will be charged to the project.

Current & Pending Support: None related to this proposal except "this proposal", $29,711. 7/01/2018-6/30/2020.

Indirect Costs

Indirect costs are 15% per the Russell Sage Foundation's requirement. Each university applied 15% to their own budget. RSF does not allow both the subcontracted organizations and the university acting as the fiscal agent to charge indirect costs on the same amount.

6 Project Output and Broader Impacts of this Research

The main outputs of this research include:

Academic presentations.

- Project 1 has been presented at the University of Michigan, Penn State University, the University of Geneva, the Upjohn institute, and the American Public Policy Association annual meetings. It is also scheduled to be presented at Iowa State University in November 2017, and the American Economic Association meetings in January 2018. We will apply to present it at the NBER 2018 Summer Institute meetings and the Society of Labor Economists 2018 annual meeting.

- We have presented Project 2 at the Carnegie-Rochester-NYU (CRNYU) and the University of Wisconsin in 2017. We plan to present it in seminars in 2018.

- We plan to present Project 3 in seminars in 2019 and 2020.

Peer-reviewed publications Project 1 currently has a working paper draft. We are preparing a manuscript of Project 1 for submission to a peer-reviewed journal. Project 2 was accepted to the Carnegie Rochester NYU Conference and as part of that conference is slated for publication after peer review to the *Journal of Monetary Economics*. We anticipate to have submit a first manuscript of Project 3 at the end of 2018.

Public use data sets and source code. One of our main objectives of this research is to create and make publicly available an occupation-year data set that can be used to study U.S. occupational change. The source code for this project will also be publicly available so that we can continue to improve the quality of the data and information extraction from the raw text. In addition, we wish to make the source code readily available for researchers who wish to apply our methodology to other newspapers. One of the broader impacts of this reseach is to generate source code that can be applied off the shelf to job vacancy data as it is collected in real-time (e.g., by EMSI or Burning Glass).

7 Project team

Enghin Atalay is an Assistant Professor at the University of Wisconsin-Madison Department of Economics. His research areas include macroeconomics and industrial organization, and he has published work in the *American Economics Review* and the *American Economic Journal: Macroeconomics.* Enghin'sresponsibilities include directing activities within the research team, leading the empirical analysis, managing project and data outputs including replication files and developing public use files, and advising and mentoring our graduate student collaborator, Phai Phongthiengtham.

Sebastian Sotelo is an Assistant Professor at the University of Michigan-Ann Arbor Department of Economics whose research is in international trade and labor markets. Sebastian's role on this project is to develop and estimate quantitative general equilibrium models of workers, tasks, and occupational sorting, and connecting these models to the reduced-form empirical analysis.

Daniel Tannenbaum an Assistant Professor in the economics department at the University of Nebraska-Lincoln whose research is in the field of empirical labor economics. Outside of this project, his work has been supported by the Spencer Foundation, and the U.S. Census Bureau and the Arnold Foundation as part of their program "Using Linked Data to Advance Evidence-Based Policymaking." Daniel's role on this project lies primarily in developing and estimating the reduced-form empirical analysis, and in writing and editing the manuscript.

All three of the PIs hold Ph.D.s from the University of Chicago in economics, and began this project in their final year as graduate students.

References

Acemoglu, D., and D. Autor (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Hanbook of Labor Economics*, 4, 1043–1171.

Akerman, A., I. Gaarder, and M. Mogstad (2015): "The Skill Complementarity of Broadband Internet," *Quarterly Journal of Economics*, 130(4), 1781–1824.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2017): "The Evolving U.S. Occupational Structure," Discussion paper.

Autor, D. H. (2013): "The "Task Approach" to Labor Markets: An Overview," *Journal of Labor Market Research*, 46(3), 185 – 199.

Autor, D. H., and D. Dorn (2013): "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103(5), 1553–97.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103(6), 2121–68.

Autor, D. H., L. Katz, and M. Kearney (2005): "Rising Wage Inequality: The Role of Composition and Prices," Discussion paper.

Autor, D. H., L. Katz, and A. B. Krueger (1998): "Computing Inequality: Have Computers Changed the Labor Market?," *Quarterly Journal of Economics*, 113(4), 1169–1213.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, 118(4), 1279–1333.

Bartel, A., C. Ichniowski, and K. Shaw (2007): "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills," *The Quarterly Journal of Economics*, 122(4), 1721–1758.

Berman, E., J. Bound, and S. Machin (1998): "Implications of Skill-Biased Technological Change: International Evidence," *Quarterly Journal of Economics*, 113(4), 1245–1280.

Brynjolfsson, E., and L. M. Hitt (2003): "Computing Productivity: Firm-Level Evidence," *Review of Economics and Statistics*, 85(4), 793–808.

Burstein, A., E. Morales, and J. Vogel (2015): "Accounting for Changes in Between-Group Inequality," Working Paper 20855, National Bureau of Economic Research.

Deming, D., and L. B. Kahn (2017): "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals," *Journal of Labor Economics*, (forthcoming).

Deming, D. J. (2017): "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, (forthcoming).

Entorf, H., and F. Kramarz (1998): "The Impact Of New Technologies On Wages: Lessons From Matching Panels On Employees And On Their Firms," *Economics of Innovation and New Technology*, 5(2-4), 165–198.

Firpo, S., N. M. Fortin, and T. Lemieux (2014): "Occupational Tasks and Changes in the Wage Structure," IZA Discussion Papers 5542, Institute for the Study of Labor (IZA).

Goos, M., and A. Manning (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *Review of Economics and Statistics*, 89(1), 118–133.

Gort, M., and S. Klepper (1982): "Time Paths in the Diffusion of Product Innovations," *Economic Journal*, 92(367), 630–653.

Griliches, Z. (1957): "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 25(4), 501–522.

Haisken-DeNew, J., and C. Schmidt (1999): "Money for Nothing and Your Chips for Free? The Anatomy of the PC Wage Differential," IZA Discussion Paper No. 86.

Hershbein, B. J., and L. B. Kahn (2016): "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings," Upjohn Working Papers and Journal Articles 16-254, W.E. Upjohn Institute for Employment Research.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2016): "The Allocation of Talent and U.S. Economic Growth," Discussion paper.

Juhn, C., K. M. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 101(3), 410–442.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107(1), 35–78.

Krueger, A. B. (1993): "How Computers Have Changed the Wage Structure: Evidence From Microdata, 1984-1989," *Quarterly Journal of Economics*, 108(1), 33–61.

Krusell, P., L. E. Ohanian, J.-V. Rios-Rull, and G. L. Violante (2000): "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 68(5), 1029–1053.

Marinescu, I., and R. Wolthoff (2016): "Opening the Black Box of the Matching Function: the Power of Words," NBER Working Papers 22508, National Bureau of Economic Research, Inc.

Michaels, G., F. Rauch, and S. J. Redding (2016): "Tasks and Technology in the United States 1880-2000," Discussion paper, National Bureau of Economic Research.

Modestino, A. S., D. Shoag, and J. Ballance (2016): "Upskilling: Do Employers Demand Greater Skill When Skilled Workers are Plentiful?," Unpublished working paper.

National Research Council (1999): *The Changing Nature of Work: Implications for Occupational Analysis*. National Academy Press.

Nordhaus, W. D. (2007): "Two Centuries of Productivity Growth in Computing," *Journal of Economic History*, 67(1), 128–159.

Spitz-Oener, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure," *Journal of Labor Economics*, 24(2), 235–270.

Figures and Tables

Table 1: Common Occupations

Job Title		6-Digit SOC Occupations		4-Digit SOC Occupations	
Description	Count	Description	Count	Description	Count
Secretary	117.3	436014: Secretary	275.6	4360: Secretary	408.6
Sales	55.4	439061: Office Clerk	182.2	4390: Oth. Admin. Support	340.3
Assistant	53.6	132011: Accountant	153.9	1320: Accountant	194.6
Accounting	51.0	412031: Retail Sales	149.8	1511: Computer Sci.	185.4
Clerk	50.0	433031: Bookkeeper	131.0	4330: Financial Clerks	174.8
Typist	50.0	439022: Typist	118.1	4120: Retail Sales	172.0
Salesperson	42.0	291141: Nurse	117.8	2911: Nurse	140.8
Engineer	41.6	111021: General Mgr.	91.6	1721: Engineers	132.1
Manager	41.3	172112: Ind. Engineer	88.1	1110: General Mgr.	127.7
Bookkeeper	40.5	113031: Financial Mgr.	68.1	1130: Financial Mgr.	101.8

Notes: This table lists the top ten job titles (columns 1-2), the top 10 6-digit SOC codes (columns 3-4), and the top 10 4-digit SOC codes (columns 5-6) in the *Boston Globe*, *New York Times*, and *Wall Street Journal* data. The counts are given in thousands of newspaper job ads.

Figure 1: Processing Newspaper Job Ads

Display Ad 133 -- No Title
Boston Globe (1960-1985); Nov 4, 1979; ProQuest Historical Newspapers: pg. E51

BUSINESS HELP

Grow with us at Fidelity!

Call 367-8960
Sunday, 3PM-7PM

or Monday thru Friday, 9AM-5PM

Fidelity is a leading financial services organization which is experiencing rapid growth through continued innovation and diversification. If you are a highly motivated person who takes pride in your work and the company you work for, consider a career with Fidelity.

MUTUAL FUNDS CLERKS
Individuals with 1-2 years' mutual funds transfer experience to process Keogh/IRA accounts and adjustments.

PAYMENTS CLERKS
Varied responsibilities processing new account applications and payments.

SHAREHOLDER SERVICES CLERKS
Interesting positions offering diversified responsibilities, processing and maintaining client accounts.

PERSONNEL CLERK
Interesting and diversified position for a well organized individual with strong record-keeping abilities. To process personnel changes and maintain insurance records. Typing 50 wpm.

ACCOUNTING CLERKS
With A/R, A/P experience. To reconcile account activity in our Data Control Department. Experience with computerized systems a plus.

STATISTICAL TYPIST
To work in our Word Processing Department. Typing: 55 wpm minimum.

MAIL CLERKS
Monday thru Friday, 12:30 AM-8:30 AM
Diversified responsibilities in our General Services Department involving sorting and distribution of mail, messenger trips, mail processing.

SECRETARIES: Investments, Sales, Brokerage
Interesting opportunities in dynamic and rapidly growing environments with diversified responsibilities. Typing 55 wpm.

Fidelity provides an outstanding fringe benefits package for our employees. We are located in downtown Boston, convenient to the MBTA, Government Center and Quincy Market.

We encourage you to grow with Fidelity and share in our success.

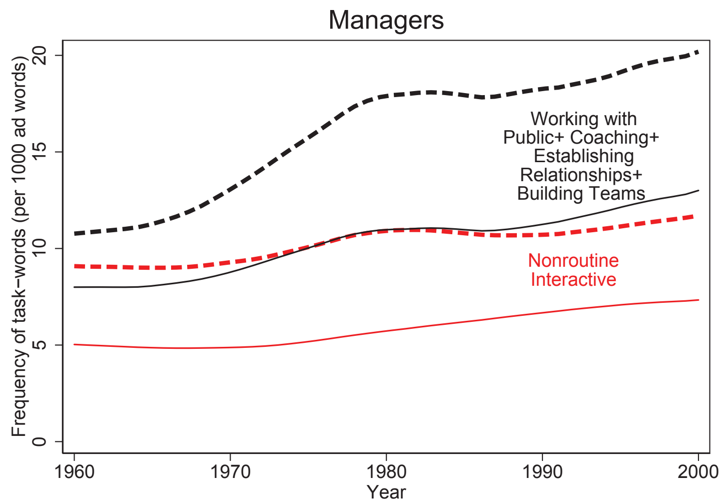
an equal opportunity employer M/F

Fidelity Group
82 DEVONSHIRE STREET

```
rapid growth through continued innovation and
diversification If you are highly motivated person who takes
pride in your work and the company you work for consider
career with Fidelity \n MUTUAL FUNDS CLERKS \n
Individuals with 1-2 years funds transfer experience to
process Keogh IRA accounts and adjustments \n PAYMENTS
CLERKS \n Varied les processing new account applications and
payments \n StI nt flt ner fr nlr ct tfl \n Mnirk \n and
maintaining client \n strong record-keeping 50 wpm \n Data
Control Department \n sorting and \n Brokerage \n
environments with \n benefits package for our Boston
convenient to the Market \n gr WilnliE lU5lty \n success \n
-l 1?Q1.a1 ol P-1Sl1v MPloV \n Fidelit \n Group \n
82 DEVONSHIRE STREET BOSTON MA 02109 \n 111 \n -l \n
TERMINAL OPERATOR \n
Position involves typing policy related information Into
computer terminal No previous computer experience required
Typing 5055 wpm Excellent benefits plus work Incentive
program in addition to starting salary of S150-165.
```

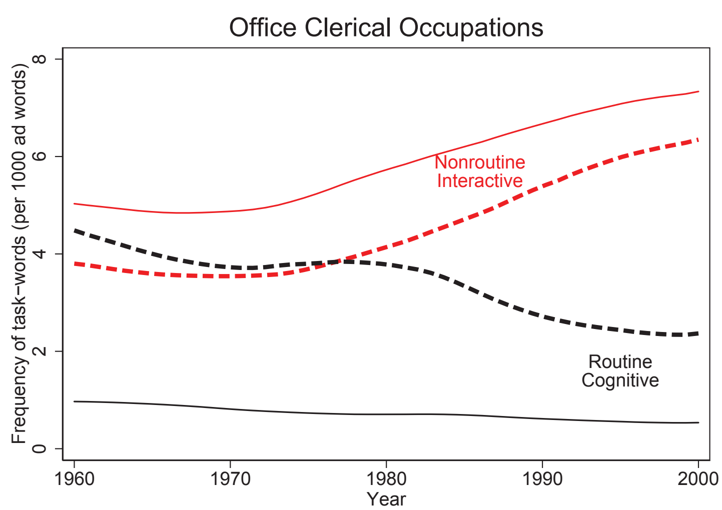
```
rapid growth through continued innovation and diversification If you are
highly motivated person who takes pride in your work and the company you
work for consider career with Fidelity \n
--------------------------------------------------------------
MUTUAL FUNDS CLERKS \n
Individuals with 1-2 years funds transfer experience to process Keogh IRA
accounts and adjustments \n
--------------------------------------------------------------
PAYMENTS CLERKS \n
Varied les processing new account applications and payments \n StI nt flt
ner fr nlr ct tfl \n Mnirk \n and maintaining client \n strong record-
keeping 50 wpm \n Data Control Department \n sorting and \n Brokerage \n
environments with \n benefits package for our Boston convenient to the
Market \n gr WilnliE lU5lty \n success \n -l 1?Q1.a1 ol P-1Sl1v MPloV \n
Fidelit \n Group \n
82 DEVONSHIRE STREET BOSTON MA 02109 \n
--------------------------------------------------------------
111 \n -l \n
--------------------------------------------------------------
TERMINAL OPERATOR \n
Position involves typing policy related information Into computer
terminal No previous computer experience required Typing 5055 wpm
Excellent benefits plus work Incentive program in addition to
starting salary of S150-165.
```

Figure 2: Task Measures: Managers and All Occupations



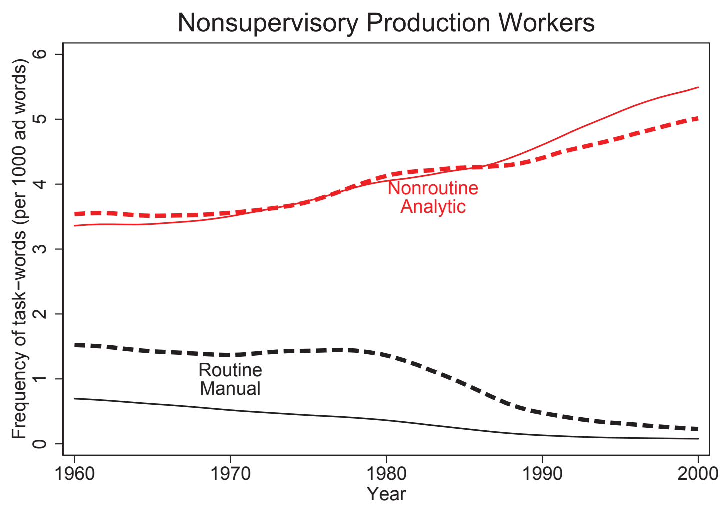
Notes: The figure above plots the average number of task-related word mentions for two different types of tasks, both for managerial occupations (dashed lines) and for all occupations (solid lines). We apply a local polynomial smoother. Managerial occupations are defined as those with a SOC code between 1100 and 1199.

Figure 3: Task Measures: Office Clerks and All Occupations



Notes: Office clerical occupations are those which have a SOC code between 4330 and 4341. We apply a local polynomial smoother. Averages across all occupations are depicted as the solid lines, and those across office clerical occupations are plotted as dashed lines.

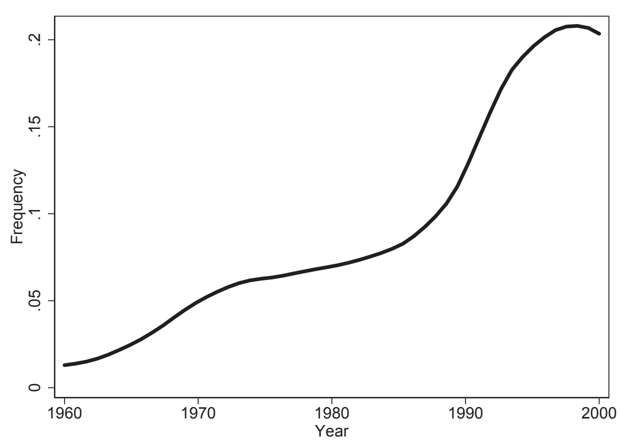
Figure 4: Task Measures: Non-Supervisory Production Workers and All Occupations



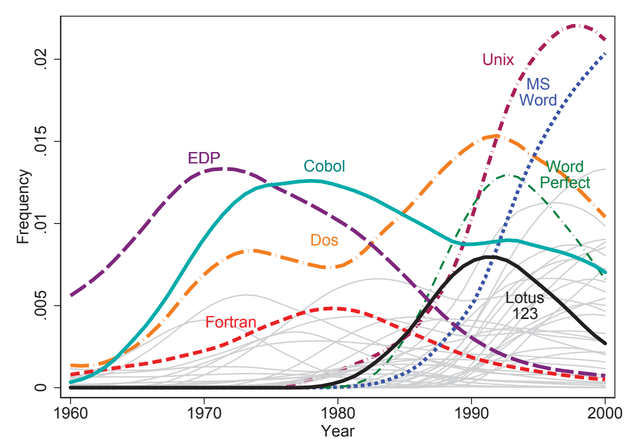
Notes: Non-supervisory production occupations have an SOC code between 5120 and 5199. We apply a local polynomial smoother. Averages across all occupations are depicted as solid lines, and those across non-supervisory production occupations are plotted as dashed lines.

Figure 5: Mentions of Technologies

(a) Total

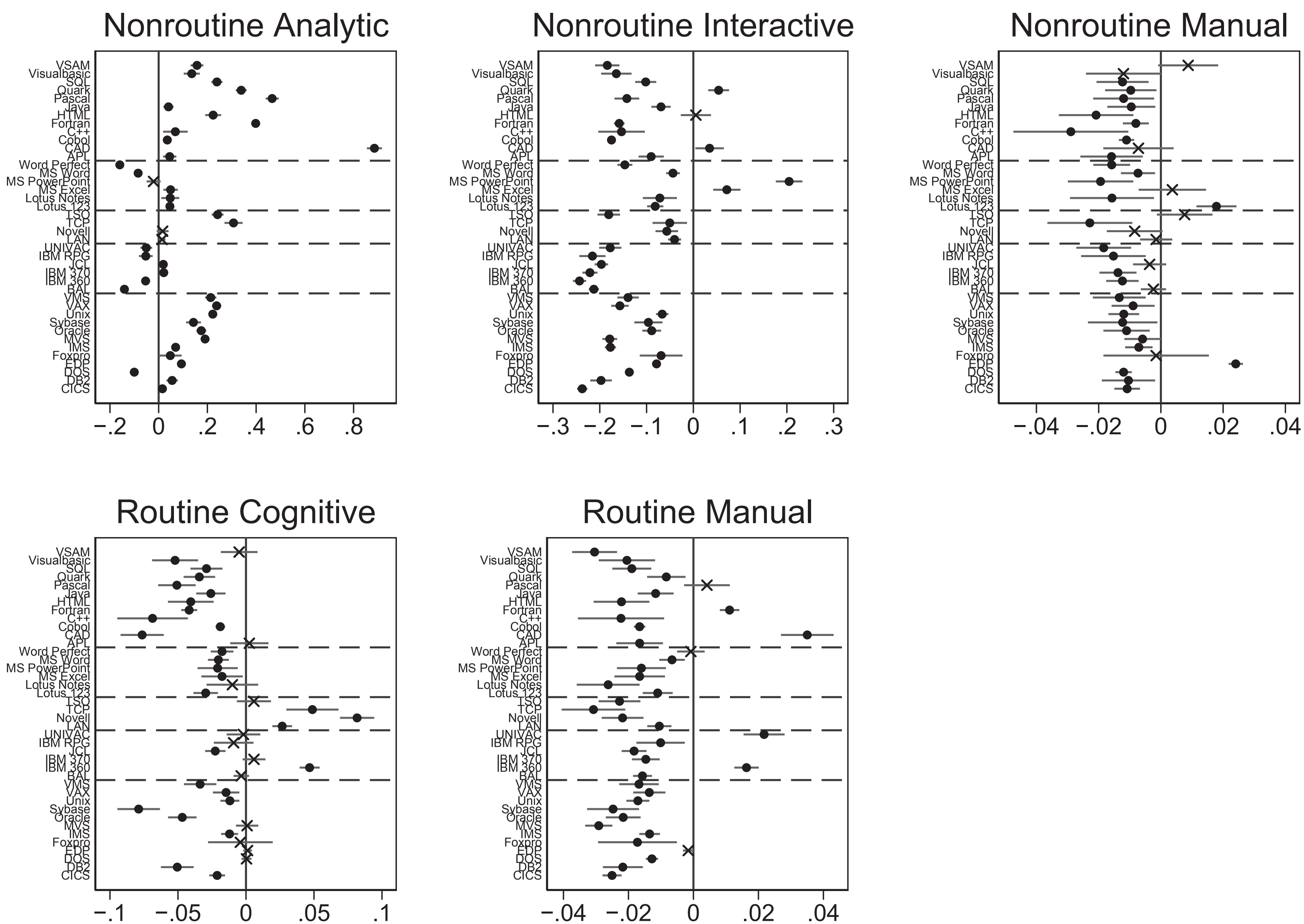


(b) By Technology



Notes: This plot gives the smoothed frequency with which job ads mention our set of technologies. The left panel depicts the sum frequency of all 40 technologies. The right panel depicts the frequencies of each of the 40 technologies separately, eight which are highlighted in thick dark lines and thirty-two which are depicted by thin, light gray lines.

Figure 6: Relationship between Task and Technology Mentions



Notes: Each panel presents the 40 coefficient estimates and corresponding 2-standard deviation confidence intervals, one for each technology, of β_{hk} from Equation 1. An "•" indicates that the coefficient estimate significantly differs from zero, while an "×" indicates that the coefficient estimate does not. Horizontal, dashed lines separate technologies into the following groups: general software, office software, networking software/hardware, other hardware, and database management systems.

Table 2: Top Occupations by Spitz-Oener (2006) Task Category

Nonroutine Analytic			Nonroutine Interactive		
1720: Engineers	69.3	0.89	1120: Sales Managers	70.1	1.11
1721: Engineers	132.1	0.89	4140: Sales Rep., Whole./Man.	64.1	0.80
1930: Social Scientists	5.2	0.70	4130: Sales Rep., Services	100.4	0.64
1910: Life Scientists	9.0	0.69	1110: Top Executives	127.7	0.61
1311: Business Operations	71.3	0.69	4120: Retail Sales	172.0	0.60
Nonroutine Manual			**Routine Cognitive**		
4930: Vehicle Mechanics	28.0	0.38	4330: Financial Clerks	174.8	0.22
4910: Maintenance Supervisors	14.1	0.29	4390: Other Admin. Support	340.3	0.10
4990: Other Maintenance	31.5	0.29	4320: Commun. Equip. Operators	11.0	0.09
4920: Electrical Mechanics	6.7	0.26	4310: Administrative Support	42.1	0.09
4340: Record Clerks	32.3	0.20	3730: Grounds Maintenance	1.4	0.08
Routine Manual					
5140: Metal and Plastic	32.0	0.14			
4990: Other Maintenance	31.5	0.07			
5141: Metal and Plastic	13.0	0.06			
4722: Construction Trades	2.6	0.05			
5120: Assemblers & Fabricators	9.8	0.05			

Notes: This table lists the top five 4-digit occupations according to the frequency with which different activity-related words are mentioned. Within each panel, the first column gives the SOC code and title; the second column gives the number of job ads in our data set (in thousands); and the final column gives the frequency (mentions per vacancy posting) of task-related words. Only occupations with at least 200 advertisements, representing 104 4-digit SOC codes, are included.

Table 3: Technologies

Technology	Introduction	First Year	Last Year	Frequency (%)
APL	1957	1961	1998	0.05
BAL	1964	1968	1983	0.30
CAD	1966	1981	1985	0.04
CICS	1968	1974	1998	0.30
COBOL	1959	1968	1998	0.83
C++	1983	1993	1999	0.02
DB2	1983	1989	1998	0.08
DOS	1966	1969	1999	0.72
EDP	1960	1963	1986	0.91
Fortran	1957	1965	1992	0.27
Foxpro	1989	1992	1998	0.02
HTML	1993	1996	$\geqslant$2000	0.04
IBM 360	1964	1965	1974	0.18
IBM 370	1970	1972	1982	0.13
IBM RPG	1959	1970	1992	0.04
IMS	1966	1960	$\geqslant$2000	0.26
Java	1995	1996	$\geqslant$2000	0.08
JCL	1964	1969	1998	0.17
LAN	1970	1990	1998	0.19
Lotus 123	1983	1987	1995	0.12
Lotus Notes	1989	1994	1998	0.03
MS Excel	1987	1993	$\geqslant$2000	0.04
MS PowerPoint	1990	1995	$\geqslant$2000	0.05
MS Word	1983	1993	1999	0.16
MVS	1974	1979	1998	0.15
Novell	1983	1994	1998	0.07
Oracle	1977	1995	1999	0.10
Pascal	1970	1982	1990	0.05
Quark	1987	1992	1999	0.07
SQL	1986	1993	1999	0.08
Sybase	1984	1995	1997	0.05
TCP	1974	1994	1999	0.03
TSO	1971	1977	1998	0.06
Univac	1951	1960	1984	0.06
Unix	1971	1992	1999	0.22
Vax	1977	1982	1998	0.11
VisualBasic	1991	1995	1998	0.04
VMS	1977	1985	1996	0.07
VSAM	1970	1982	1998	0.05
Word Perfect	1979	1988	1998	0.15

Notes: This table lists the 40 technologies in our sample. The "First Year" and "Last Year" columns report the first year and last year at which the frequency of technology mentions was at least one-third of the frequency of the year with the maximum mention frequency (number of technology mentions per job ad). The $\geqslant$2000 symbol indicates that the technology was still in broad use at the end of the sample period.